

# The Unicode Standard

## Version 8.0 – Core Specification

To learn about the latest version of the Unicode Standard, see <http://www.unicode.org/versions/latest/>.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and the publisher was aware of a trademark claim, the designations have been printed with initial capital letters or in all capitals.

Unicode and the Unicode Logo are registered trademarks of Unicode, Inc., in the United States and other countries.

The authors and publisher have taken care in the preparation of this specification, but make no expressed or implied warranty of any kind and assume no responsibility for errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of the use of the information or programs contained herein.

The *Unicode Character Database* and other files are provided as-is by Unicode, Inc. No claims are made as to fitness for any particular purpose. No warranties of any kind are expressed or implied. The recipient agrees to determine applicability of information provided.

Copyright © 1991–2015 Unicode, Inc.

All rights reserved. This publication is protected by copyright, and permission must be obtained from the publisher prior to any prohibited reproduction. For information regarding permissions, inquire at <http://www.unicode.org/reporting.html>. For information about the Unicode terms of use, please see <http://www.unicode.org/copyright.html>.

The Unicode Standard / the Unicode Consortium ; edited by Julie D. Allen ... [et al.]. — Version 8.0

Includes bibliographical references and index.

ISBN 978-1-936213-10-8 (<http://www.unicode.org/versions/Unicode8.0.0/>)

1. Unicode (Computer character set) I. Allen, Julie D. II. Unicode Consortium.

QA268.U545 2015

ISBN 978-1-936213-10-8

Published in Mountain View, CA

August 2015

# Contents

<b>Figures</b> .....	xxi
<b>Tables</b> .....	xxv
<b>Preface</b> .....	xxxix
Why Unicode? .....	xxxix
What's New? .....	xxxix
Organization of This Standard .....	xxxix
Unicode Standard Annexes .....	xxxix
The Unicode Character Database .....	xxxv
Unicode Code Charts .....	xxxv
Unicode Technical Standards and Unicode Technical Reports .....	xxxv
Updates and Errata .....	xxxvi
Acknowledgements .....	xxxvi
<b>1 Introduction</b> .....	1
<b>1.1 Coverage</b> .....	3
Standards Coverage .....	3
New Characters .....	3
<b>1.2 Design Goals</b> .....	4
<b>1.3 Text Handling</b> .....	6
Characters and Glyphs .....	6
Text Elements .....	6
<b>2 General Structure</b> .....	9
<b>2.1 Architectural Context</b> .....	10
Basic Text Processes .....	10
Text Elements, Characters, and Text Processes .....	10
Text Processes and Encoding .....	11
<b>2.2 Unicode Design Principles</b> .....	14
Universality .....	14
Efficiency .....	15
Characters, Not Glyphs .....	15
Semantics .....	18
Plain Text .....	18
Logical Order .....	19
Unification .....	21
Dynamic Composition .....	22
Stability .....	23
Convertibility .....	23

<b>2.3</b>	<b>Compatibility Characters</b> .....	25
	Compatibility Variants .....	26
	Compatibility Decomposable Characters .....	26
<b>2.4</b>	<b>Code Points and Characters</b> .....	29
	Types of Code Points .....	30
<b>2.5</b>	<b>Encoding Forms</b> .....	33
	UTF-32 .....	35
	UTF-16 .....	36
	UTF-8 .....	36
	Comparison of the Advantages of UTF-32, UTF-16, and UTF-8 .....	38
<b>2.6</b>	<b>Encoding Schemes</b> .....	40
<b>2.7</b>	<b>Unicode Strings</b> .....	43
<b>2.8</b>	<b>Unicode Allocation</b> .....	44
	Planes .....	44
	Allocation Areas and Character Blocks .....	45
	Assignment of Code Points .....	46
<b>2.9</b>	<b>Details of Allocation</b> .....	48
	Plane 0 (BMP) .....	49
	Plane 1 (SMP) .....	51
	Plane 2 (SIP) .....	52
	Other Planes .....	52
<b>2.10</b>	<b>Writing Direction</b> .....	53
<b>2.11</b>	<b>Combining Characters</b> .....	55
	Sequence of Base Characters and Diacritics .....	56
	Multiple Combining Characters .....	57
	Ligated Multiple Base Characters .....	59
	Exhibiting Nonspacing Marks in Isolation .....	60
	“Characters” and Grapheme Clusters .....	60
<b>2.12</b>	<b>Equivalent Sequences</b> .....	62
	Normalization .....	62
	Decompositions .....	63
	Non-decomposition of Certain Diacritics .....	65
<b>2.13</b>	<b>Special Characters</b> .....	67
	Special Noncharacter Code Points .....	67
	Byte Order Mark (BOM) .....	67
	Layout and Format Control Characters .....	68
	The Replacement Character .....	68
	Control Codes .....	68
<b>2.14</b>	<b>Conforming to the Unicode Standard</b> .....	69
	Characteristics of Conformant Implementations .....	69
	Unacceptable Behavior .....	70
	Acceptable Behavior .....	70
	Supported Subsets .....	71

<b>3</b>	<b>Conformance</b>	73
<b>3.1</b>	<b>Versions of the Unicode Standard</b>	74
	Stability	74
	Version Numbering	75
	Errata and Corrigenda	76
	References to the Unicode Standard	76
	Precision in Version Citation	77
	References to Unicode Character Properties	77
	References to Unicode Algorithms	78
<b>3.2</b>	<b>Conformance Requirements</b>	79
	Code Points Unassigned to Abstract Characters	79
	Interpretation	80
	Modification	81
	Character Encoding Forms	82
	Character Encoding Schemes	83
	Bidirectional Text	84
	Normalization Forms	84
	Normative References	84
	Unicode Algorithms	84
	Default Casing Algorithms	85
	Unicode Standard Annexes	85
<b>3.3</b>	<b>Semantics</b>	87
	Definitions	87
	Character Identity and Semantics	87
<b>3.4</b>	<b>Characters and Encoding</b>	90
<b>3.5</b>	<b>Properties</b>	94
	Types of Properties	94
	Property Values	95
	Default Property Values	96
	Classification of Properties by Their Values	97
	Property Status	98
	Context Dependence	101
	Stability of Properties	101
	Simple and Derived Properties	103
	Property Aliases	103
	Private Use	104
<b>3.6</b>	<b>Combination</b>	105
	Combining Character Sequences	105
	Grapheme Clusters	107
	Application of Combining Marks	109
<b>3.7</b>	<b>Decomposition</b>	115
	Compatibility Decomposition	115
	Canonical Decomposition	116

<b>3.8</b>	<b>Surrogates</b> .....	118
<b>3.9</b>	<b>Unicode Encoding Forms</b> .....	119
	UTF-32 .....	123
	UTF-16 .....	124
	UTF-8 .....	124
	Encoding Form Conversion .....	126
	Constraints on Conversion Processes .....	126
<b>3.10</b>	<b>Unicode Encoding Schemes</b> .....	130
<b>3.11</b>	<b>Normalization Forms</b> .....	134
	Normalization Stability .....	134
	Combining Classes .....	135
	Specification of Unicode Normalization Forms .....	136
	Starters .....	136
	Canonical Ordering Algorithm .....	137
	Canonical Composition Algorithm .....	138
	Definition of Normalization Forms .....	140
<b>3.12</b>	<b>Conjoining Jamo Behavior</b> .....	142
	Definitions .....	142
	Hangul Syllable Decomposition .....	144
	Hangul Syllable Composition .....	146
	Hangul Syllable Name Generation .....	147
	Sample Code for Hangul Algorithms .....	148
<b>3.13</b>	<b>Default Case Algorithms</b> .....	152
	Definitions .....	152
	Default Case Conversion .....	154
	Default Case Folding .....	155
	Default Case Detection .....	156
	Default Caseless Matching .....	157
<b>4</b>	<b>Character Properties</b> .....	159
<b>4.1</b>	<b>Unicode Character Database</b> .....	161
<b>4.2</b>	<b>Case</b> .....	164
	Definitions of Case and Casing .....	164
	Case Mapping .....	166
<b>4.3</b>	<b>Combining Classes</b> .....	168
	Reordrant, Split, and Subjoined Combining Marks .....	168
<b>4.4</b>	<b>Directionality</b> .....	173
<b>4.5</b>	<b>General Category</b> .....	174
<b>4.6</b>	<b>Numeric Value</b> .....	177
	Ideographic Numeric Values .....	178
<b>4.7</b>	<b>Bidi Mirrored</b> .....	180

4.8	<b>Name</b> .....	182
	Unicode Name Property .....	185
	Code Point Labels .....	186
	Use of Character Names in APIs and User Interfaces .....	187
4.9	<b>Unicode 1.0 Names</b> .....	188
4.10	<b>Letters, Alphabetic, and Ideographic</b> .....	189
4.11	<b>Properties Related to Text Boundaries</b> .....	190
4.12	<b>Characters with Unusual Properties</b> .....	191
5	<b>Implementation Guidelines</b> .....	195
5.1	<b>Data Structures for Character Conversion</b> .....	196
	Issues .....	196
	Multistage Tables .....	196
5.2	<b>Programming Languages and Data Types</b> .....	199
	Unicode Data Types for C .....	199
5.3	<b>Unknown and Missing Characters</b> .....	201
5.4	<b>Handling Surrogate Pairs in UTF-16</b> .....	203
5.5	<b>Handling Numbers</b> .....	205
5.6	<b>Normalization</b> .....	206
5.7	<b>Compression</b> .....	208
5.8	<b>Newline Guidelines</b> .....	209
	Definitions .....	209
	Line Separator and Paragraph Separator .....	210
	Recommendations .....	211
5.9	<b>Regular Expressions</b> .....	214
5.10	<b>Language Information in Plain Text</b> .....	215
	Requirements for Language Tagging .....	215
	Language Tags and Han Unification .....	215
5.11	<b>Editing and Selection</b> .....	217
	Consistent Text Elements .....	217
5.12	<b>Strategies for Handling Nonspacing Marks</b> .....	219
	Keyboard Input .....	220
	Truncation .....	220
5.13	<b>Rendering Nonspacing Marks</b> .....	222
	Canonical Equivalence .....	225
	Positioning Methods .....	226
5.14	<b>Locating Text Element Boundaries</b> .....	228
5.15	<b>Identifiers</b> .....	229
5.16	<b>Sorting and Searching</b> .....	230
	Culturally Expected Sorting and Searching .....	230
	Language-Insensitive Sorting .....	230

	Searching .....	231
	Sublinear Searching .....	231
<b>5.17</b>	<b>Binary Order</b> .....	233
	UTF-8 in UTF-16 Order .....	233
	UTF-16 in UTF-8 Order .....	234
<b>5.18</b>	<b>Case Mappings</b> .....	236
	Titlecasing .....	236
	Complications for Case Mapping .....	237
	Reversibility .....	239
	Caseless Matching .....	240
	Normalization and Casing .....	242
<b>5.19</b>	<b>Mapping Compatibility Variants</b> .....	244
<b>5.20</b>	<b>Unicode Security</b> .....	246
<b>5.21</b>	<b>Ignoring Characters in Processing</b> .....	249
	Characters Ignored in Text Segmentation .....	249
	Characters Ignored in Line Breaking .....	250
	Characters Ignored in Cursive Joining .....	250
	Characters Ignored in Identifiers .....	250
	Characters Ignored in Searching and Sorting .....	251
	Characters Ignored for Display .....	251
<b>5.22</b>	<b>Best Practice for U+FFFD Substitution</b> .....	255
<b>6</b>	<b>Writing Systems and Punctuation</b> .....	257
<b>6.1</b>	<b>Writing Systems</b> .....	258
<b>6.2</b>	<b>General Punctuation</b> .....	263
	Blocks Devoted to Punctuation .....	264
	Format Control Characters .....	265
	Space Characters .....	266
	Dashes and Hyphens .....	267
	Paired Punctuation .....	269
	Language-Based Usage of Quotation Marks .....	270
	Apostrophes .....	274
	Other Punctuation .....	274
	Archaic Punctuation and Editorial Marks .....	279
	Indic Punctuation .....	282
	CJK Punctuation .....	284
	Unknown or Unavailable Ideographs .....	285
	CJK Compatibility Forms .....	286
<b>7</b>	<b>Europe-I</b> .....	289
	<b>Modern and Liturgical Scripts</b>	
<b>7.1</b>	<b>Latin</b> .....	291
	Letters of Basic Latin: U+0041–U+007A .....	295
	Letters of the Latin-1 Supplement: U+00C0–U+00FF .....	295
	Latin Extended-A: U+0100–U+017F .....	295

	Latin Extended-B: U+0180–U+024F	296
	IPA Extensions: U+0250–U+02AF	298
	Phonetic Extensions: U+1D00–U+1DBF	300
	Latin Extended Additional: U+1E00–U+1EFF	301
	Latin Extended-C: U+2C60–U+2C7F	301
	Latin Extended-D: U+A720–U+A7FF	302
	Latin Extended-E: U+AB30–U+AB6F	303
	Latin Ligatures: U+FB00–U+FB06	303
<b>7.2</b>	<b>Greek</b>	304
	Greek: U+0370–U+03FF	304
	Greek Extended: U+1F00–U+1FFF	308
	Ancient Greek Numbers: U+10140–U+1018F	309
<b>7.3</b>	<b>Coptic</b>	311
<b>7.4</b>	<b>Cyrillic</b>	314
	Cyrillic: U+0400–U+04FF	314
	Cyrillic Supplement: U+0500–U+052F	315
	Cyrillic Extended-A: U+2DE0–U+2DFF	315
	Cyrillic Extended-B: U+A640–U+A69F	316
<b>7.5</b>	<b>Glagolitic</b>	318
<b>7.6</b>	<b>Armenian</b>	319
<b>7.7</b>	<b>Georgian</b>	321
<b>7.8</b>	<b>Modifier Letters</b>	323
	Spacing Modifier Letters: U+02B0–U+02FF	324
	Modifier Tone Letters: U+A700–U+A71F	326
<b>7.9</b>	<b>Combining Marks</b>	327
	Combining Diacritical Marks: U+0300–U+036F	332
	Combining Diacritical Marks Extended: U+1AB0–U+1AFF	333
	Combining Diacritical Marks Supplement: U+1DC0–U+1DFF	333
	Combining Marks for Symbols: U+20D0–U+20FF	334
	Combining Half Marks: U+FE20–U+FE2F	335
	Combining Marks in Other Blocks	335
<b>8</b>	<b>Europe-II</b>	337
	<b>Ancient and Other Scripts</b>	
<b>8.1</b>	<b>Linear A</b>	339
<b>8.2</b>	<b>Linear B</b>	340
	Linear B Syllabary: U+10000–U+1007F	340
	Linear B Ideograms: U+10080–U+100FF	340
	Aegean Numbers: U+10100–U+1013F	340
<b>8.3</b>	<b>Cypriot Syllabary</b>	342



<b>8.4</b>	<b>Ancient Anatolian Alphabets</b> .....	343
	Lycian: U+10280–U+1029F .....	343
	Carian: U+102A0–U+102DF .....	343
	Lydian: U+10920–U+1093F .....	343
<b>8.5</b>	<b>Old Italic</b> .....	345
<b>8.6</b>	<b>Runic</b> .....	348
<b>8.7</b>	<b>Old Hungarian</b> .....	351
<b>8.8</b>	<b>Gothic</b> .....	352
<b>8.9</b>	<b>Elbasan</b> .....	353
<b>8.10</b>	<b>Caucasian Albanian</b> .....	354
<b>8.11</b>	<b>Old Permic</b> .....	355
<b>8.12</b>	<b>Ogham</b> .....	356
<b>8.13</b>	<b>Shavian</b> .....	357
<b>9</b>	<b>Middle East-I</b> .....	359
	<b>Modern and Liturgical Scripts</b>	
<b>9.1</b>	<b>Hebrew</b> .....	361
	Hebrew: U+0590–U+05FF .....	361
	Alphabetic Presentation Forms: U+FB1D–U+FB4F .....	365
<b>9.2</b>	<b>Arabic</b> .....	367
	Arabic: U+0600–U+06FF .....	367
	Arabic Cursive Joining .....	374
	Arabic Ligatures .....	377
	Arabic Joining Groups .....	379
	Arabic Supplement: U+0750–U+077F .....	386
	Arabic Extended-A: U+08A0–U+08FF .....	386
	Arabic Presentation Forms-A: U+FB50–U+FDFF .....	386
	Arabic Presentation Forms-B: U+FE70–U+FEFF .....	388
<b>9.3</b>	<b>Syriac</b> .....	389
	Syriac: U+0700–U+074F .....	389
	Syriac Shaping .....	393
<b>9.4</b>	<b>Samaritan</b> .....	398
<b>9.5</b>	<b>Mandaic</b> .....	400
<b>10</b>	<b>Middle East-II</b> .....	403
	<b>Ancient Scripts</b>	
<b>10.1</b>	<b>Old North Arabian</b> .....	405
<b>10.2</b>	<b>Old South Arabian</b> .....	406
<b>10.3</b>	<b>Phoenician</b> .....	408
<b>10.4</b>	<b>Imperial Aramaic</b> .....	410
<b>10.5</b>	<b>Manichaean</b> .....	412

<b>10.6 Pahlavi and Parthian</b> .....	416
Inscriptional Parthian: U+10B40–U+10B5F .....	416
Inscriptional Pahlavi: U+10B60–U+10B7F .....	416
Psalter Pahlavi: U+10B80–U+10BAF .....	417
<b>10.7 Avestan</b> .....	418
<b>10.8 Nabataean</b> .....	420
<b>10.9 Palmyrene</b> .....	421
<b>10.10 Hatran</b> .....	422
<b>11 Cuneiform and Hieroglyphs</b> .....	423
<b>11.1 Sumero-Akkadian</b> .....	424
Cuneiform: U+12000–U+123FF .....	424
Cuneiform Numbers and Punctuation: U+12400–U+1247F .....	427
Early Dynastic Cuneiform: U+12480–U+1254F .....	427
<b>11.2 Ugaritic</b> .....	428
<b>11.3 Old Persian</b> .....	429
<b>11.4 Egyptian Hieroglyphs</b> .....	430
<b>11.5 Meroitic</b> .....	434
<b>11.6 Anatolian Hieroglyphs</b> .....	436
<b>12 South and Central Asia-I</b> .....	439
<b>Official Scripts of India</b>	
<b>12.1 Devanagari</b> .....	441
Devanagari: U+0900–U+097F .....	441
Principles of the Devanagari Script .....	442
Rendering Devanagari .....	448
Devanagari Digits, Punctuation, and Symbols .....	457
Extensions in the Main Devanagari Block .....	458
Devanagari Extended: U+A8E0–U+A8FF .....	460
Vedic Extensions: U+1CD0–U+1CFF .....	461
<b>12.2 Bengali (Bangla)</b> .....	463
<b>12.3 Gurmukhi</b> .....	469
<b>12.4 Gujarati</b> .....	474
<b>12.5 Oriya (Odia)</b> .....	476
<b>12.6 Tamil</b> .....	479
Tamil: U+0B80–U+0BFF .....	479
Tamil Vowels .....	480
Tamil Ligatures .....	481
Tamil Named Character Sequences .....	485
<b>12.7 Telugu</b> .....	488
<b>12.8 Kannada</b> .....	491
Kannada: U+0C80–U+0CFF .....	491

	Principles of the Kannada Script .....	491
	Rendering Kannada .....	493
<b>12.9</b>	<b>Malayalam</b> .....	495
	Malayalam Orthographic Reform .....	495
	Rendering Malayalam .....	496
	Malayalam Numbers and Punctuation .....	501
<b>13</b>	<b>South and Central Asia-II</b> .....	503
	<b>Other Modern Scripts</b>	
<b>13.1</b>	<b>Thaana</b> .....	505
<b>13.2</b>	<b>Sinhala</b> .....	507
	Sinhala Archaic Numbers: U+111E0–U+111FF .....	508
<b>13.3</b>	<b>Tibetan</b> .....	509
<b>13.4</b>	<b>Mongolian</b> .....	521
<b>13.5</b>	<b>Limbu</b> .....	530
<b>13.6</b>	<b>Meetei Mayek</b> .....	534
<b>13.7</b>	<b>Mro</b> .....	536
<b>13.8</b>	<b>Warang Citi</b> .....	537
<b>13.9</b>	<b>Ol Chiki</b> .....	538
<b>13.10</b>	<b>Chakma</b> .....	540
<b>13.11</b>	<b>Lepcha</b> .....	541
<b>13.12</b>	<b>Saurashtra</b> .....	544
<b>14</b>	<b>South and Central Asia-III</b> .....	545
	<b>Ancient Scripts</b>	
<b>14.1</b>	<b>Brahmi</b> .....	546
<b>14.2</b>	<b>Kharoshthi</b> .....	550
	Kharoshthi: U+10A00–U+10A5F .....	550
	Rendering Kharoshthi .....	552
<b>14.3</b>	<b>Phags-pa</b> .....	556
<b>14.4</b>	<b>Old Turkic</b> .....	563
<b>15</b>	<b>South and Central Asia-IV</b> .....	565
	<b>Other Historic Scripts</b>	
<b>15.1</b>	<b>Syoti Nagri</b> .....	567
<b>15.2</b>	<b>Kaithi</b> .....	569
<b>15.3</b>	<b>Sharada</b> .....	572
<b>15.4</b>	<b>Takri</b> .....	574
<b>15.5</b>	<b>Siddham</b> .....	576
<b>15.6</b>	<b>Mahajani</b> .....	578
<b>15.7</b>	<b>Khojki</b> .....	580
<b>15.8</b>	<b>Khudawadi</b> .....	582

15.9	<b>Multani</b> .....	584
15.10	<b>Tirhuta</b> .....	585
15.11	<b>Modi</b> .....	588
15.12	<b>Grantha</b> .....	591
	Rendering Behavior .....	591
15.13	<b>Ahom</b> .....	594
15.14	<b>Sora Sompeng</b> .....	596
16	<b>Southeast Asia</b> .....	597
16.1	<b>Thai</b> .....	599
16.2	<b>Lao</b> .....	603
16.3	<b>Myanmar</b> .....	606
	Myanmar: U+1000–U+109F .....	606
	Myanmar Extended-A: U+AA60–U+AA7F .....	610
	Khamti Shan .....	610
	Aiton and Phake .....	612
	Myanmar Extended-B: U+A9E0–U+A9FF .....	612
16.4	<b>Khmer</b> .....	613
	Khmer: U+1780–U+17FF .....	613
	Principles of the Khmer Script .....	613
	Khmer Symbols: U+19E0–U+19FF .....	623
16.5	<b>Tai Le</b> .....	624
16.6	<b>New Tai Lue</b> .....	626
16.7	<b>Tai Tham</b> .....	629
16.8	<b>Tai Viet</b> .....	632
16.9	<b>Kayah Li</b> .....	635
16.10	<b>Cham</b> .....	636
16.11	<b>Pahawh Hmong</b> .....	638
16.12	<b>Pau Cin Hau</b> .....	640
17	<b>Indonesia and Oceania</b> .....	641
17.1	<b>Philippine Scripts</b> .....	642
	Tagalog: U+1700–U+171F .....	642
	Hanunóo: U+1720–U+173F .....	642
	Buhid: U+1740–U+175F .....	642
	Tagbanwa: U+1760–U+177F .....	642
	Principles of the Philippine Scripts .....	642
17.2	<b>Buginese</b> .....	645
17.3	<b>Balinese</b> .....	647
17.4	<b>Javanese</b> .....	653
17.5	<b>Rejang</b> .....	657
17.6	<b>Batak</b> .....	658

<b>17.7 Sundanese</b> .....	659
Sundanese: U+1B80–U+1BBF .....	659
Sundanese Supplement: U+1CC0–U+1CCF .....	660
<b>18 East Asia</b> .....	661
<b>18.1 Han</b> .....	663
CJK Unified Ideographs .....	663
Blocks Containing Han Ideographs .....	664
General Characteristics of Han Ideographs .....	666
Principles of Han Unification .....	670
Unification Rules .....	671
Abstract Shape .....	672
Han Ideograph Arrangement .....	674
Radical-Stroke Indices .....	675
Mappings for Han Ideographs .....	676
CJK Unified Ideographs Extension B: U+2000–U+2A6D6 .....	677
CJK Unified Ideographs Extension C: U+2A700–U+2B734 .....	677
CJK Unified Ideographs Extension D: U+2B740–U+2B81D .....	677
CJK Unified Ideographs Extension E: U+2B820–U+2CEA1 .....	677
CJK Compatibility Ideographs: U+F900–U+FAFF .....	677
CJK Compatibility Supplement: U+2F800–U+2FA1D .....	678
Kanbun: U+3190–U+319F .....	678
Symbols Derived from Han Ideographs .....	679
CJK and KangXi Radicals: U+2E80–U+2FD5 .....	679
CJK Additions from HKSCS and GB 18030 .....	680
CJK Strokes: U+31C0–U+31EF .....	680
<b>18.2 Ideographic Description Characters</b> .....	681
<b>18.3 Bopomofo</b> .....	685
<b>18.4 Hiragana and Katakana</b> .....	688
Hiragana: U+3040–U+309F .....	688
Katakana: U+30A0–U+30FF .....	688
Katakana Phonetic Extensions: U+31F0–U+31FF .....	689
Kana Supplement U+1B000–U+1BOFF .....	689
<b>18.5 Halfwidth and Fullwidth Forms</b> .....	690
<b>18.6 Hangul</b> .....	691
Hangul Jamo: U+1100–U+11FF .....	691
Hangul Jamo Extended-A: U+A960–U+A97F .....	692
Hangul Jamo Extended-B: U+D7B0–U+D7FF .....	692
Hangul Compatibility Jamo: U+3130–U+318F .....	692
Hangul Syllables: U+AC00–U+D7A3 .....	693
<b>18.7 Yi</b> .....	695
<b>18.8 Lisu</b> .....	698
<b>18.9 Miao</b> .....	701

<b>19 Africa</b> .....	703
<b>19.1 Ethiopic</b> .....	704
Ethiopic: U+1200–U+137F .....	704
Ethiopic Extensions .....	706
<b>19.2 Osmanya</b> .....	708
<b>19.3 Tifinagh</b> .....	709
<b>19.4 N’Ko</b> .....	712
<b>19.5 Vai</b> .....	717
<b>19.6 Bamum</b> .....	719
Bamum: U+A6A0–U+A6FF .....	719
Bamum Supplement: U+16800–U+16A3F .....	719
<b>19.7 Bassa Vah</b> .....	721
<b>19.8 Mende Kikakui</b> .....	722
<b>20 Americas</b> .....	725
<b>20.1 Cherokee</b> .....	726
<b>20.2 Canadian Aboriginal Syllabics</b> .....	728
Canadian Aboriginal Syllabics: U+1400–U+167F .....	728
Canadian Aboriginal Syllabics Extended: U+18B0–U+18FF .....	729
<b>20.3 Deseret</b> .....	730
<b>21 Notational Systems</b> .....	733
<b>21.1 Braille</b> .....	734
<b>21.2 Western Musical Symbols</b> .....	736
<b>21.3 Byzantine Musical Symbols</b> .....	741
<b>21.4 Ancient Greek Musical Notation</b> .....	742
<b>21.5 Duployan</b> .....	745
<b>21.6 Sutton SignWriting</b> .....	747
<b>22 Symbols</b> .....	749
<b>22.1 Currency Symbols</b> .....	751
<b>22.2 Letterlike Symbols</b> .....	754
Letterlike Symbols: U+2100–U+214F .....	754
Mathematical Alphanumeric Symbols: U+1D400–U+1D7FF .....	755
Mathematical Alphabets .....	756
Fonts Used for Mathematical Alphabets .....	758
Arabic Mathematical Alphabetic Symbols: U+1EE00–U+1EEFF .....	760
<b>22.3 Numerals</b> .....	761
Decimal Digits .....	761
Other Digits .....	763
Non-Decimal Radix Systems .....	765
Acrophonic Systems and Other Letter-based Numbers .....	766
Coptic Epact Numbers: U+102E0–U+102FF .....	767
Rumi Numeral Forms: U+10E60–U+10E7E .....	768

	CJK Numerals	768
	Fractions	770
	Common Indic Number Forms: U+A830–U+A83F	770
<b>22.4</b>	<b>Superscript and Subscript Symbols</b>	772
	Superscripts and Subscripts: U+2070–U+209F	772
<b>22.5</b>	<b>Mathematical Symbols</b>	774
	Mathematical Operators: U+2200–U+22FF	775
	Supplements to Mathematical Symbols and Arrows	778
	Supplemental Mathematical Operators: U+2A00–U+2AFF	778
	Miscellaneous Mathematical Symbols-A: U+27C0–U+27EF	778
	Miscellaneous Mathematical Symbols-B: U+2980–U+29FF	779
	Miscellaneous Symbols and Arrows: U+2B00–U+2B7F	779
	Arrows: U+2190–U+21FF	779
	Supplemental Arrows	780
	Standardized Variants of Mathematical Symbols	781
<b>22.6</b>	<b>Invisible Mathematical Operators</b>	782
<b>22.7</b>	<b>Technical Symbols</b>	783
	Control Pictures: U+2400–U+243F	783
	Miscellaneous Technical: U+2300–U+23FF	783
	Optical Character Recognition: U+2440–U+245F	787
<b>22.8</b>	<b>Geometrical Symbols</b>	788
	Box Drawing and Block Elements	788
	Geometric Shapes: U+25A0–U+25FF	789
	Geometric Shapes Extended: U+1F780–U+1F7FF	790
<b>22.9</b>	<b>Miscellaneous Symbols</b>	791
	Miscellaneous Symbols and Pictographs	792
	Emoticons: U+1F600–U+1F64F	795
	Transport and Map Symbols: U+1F680–U+1F6FF	795
	Dingbats: U+2700–U+27BF	795
	Ornamental Dingbats: U+1F650–U+1F67F	797
	Alchemical Symbols: U+1F700–U+1F77F	797
	Mahjong Tiles: U+1F000–U+1F02F	798
	Domino Tiles: U+1F030–U+1F09F	798
	Playing Cards: U+1F0A0–U+1F0FF	798
	Yijing Hexagram Symbols: U+4DC0–U+4DFF	799
	Tai Xuan Jing Symbols: U+1D300–U+1D356	800
	Ancient Symbols: U+10190–U+101CF	800
	Phaistos Disc Symbols: U+101D0–U+101FF	801
<b>22.10</b>	<b>Enclosed and Square</b>	802
	Enclosed Alphanumerics: U+2460–U+24FF	804
	Enclosed CJK Letters and Months: U+3200–U+32FF	804
	CJK Compatibility: U+3300–U+33FF	805
	Enclosed Alphanumeric Supplement: U+1F100–U+1F1FF	805
	Enclosed Ideographic Supplement: U+1F200–U+1F2FF	806

<b>23</b>	<b>Special Areas and Format Characters</b>	807
<b>23.1</b>	<b>Control Codes</b>	808
	Representing Control Sequences	808
	Specification of Control Code Semantics	809
<b>23.2</b>	<b>Layout Controls</b>	811
	Line and Word Breaking	811
	Cursive Connection and Ligatures	813
	Combining Grapheme Joiner	817
	Bidirectional Ordering Controls	819
	Stateful Format Controls	820
<b>23.3</b>	<b>Deprecated Format Characters</b>	822
<b>23.4</b>	<b>Variation Selectors</b>	824
<b>23.5</b>	<b>Private-Use Characters</b>	827
	Private Use Area: U+E000–U+F8FF	828
	Supplementary Private Use Areas	829
<b>23.6</b>	<b>Surrogates Area</b>	830
<b>23.7</b>	<b>Noncharacters</b>	831
<b>23.8</b>	<b>Specials</b>	833
	Byte Order Mark (BOM): U+FEFF	833
	Specials: U+FFF0–U+FFF8	835
	Annotation Characters: U+FFF9–U+FFFB	835
	Replacement Characters: U+FFFC–U+FFFD	837
<b>23.9</b>	<b>Tag Characters</b>	838
	Tag Characters: U+E000–U+E007F	838
	Deprecated Use for Language Tagging	838
	Syntax for Embedding Tags	838
	Working with Language Tags	840
	Unicode Conformance Issues	842
	Formal Tag Syntax	842
<b>24</b>	<b>About the Code Charts</b>	845
<b>24.1</b>	<b>Character Names List</b>	846
	Images in the Code Charts and Character Lists	846
	Special Characters and Code Points	848
	Character Names	850
	Informative Aliases	850
	Normative Aliases	850
	Cross References	850
	Information About Languages	851
	Case Mappings	851
	Decompositions	852
	Standardized Variation Sequences	853
	Subheads	854



24.2	<b>CJK Ideographs</b> .....	855
	CJK Unified Ideographs .....	855
	Compatibility Ideographs .....	856
24.3	<b>Hangul Syllables</b> .....	858
<b>A</b>	<b>Notational Conventions</b> .....	859
	Code Points .....	859
	Character Names .....	859
	Character Blocks .....	859
	Sequences .....	860
	Rendering .....	860
	Properties and Property Values .....	861
	Miscellaneous .....	861
	Extended BNF .....	861
	Operators .....	863
<b>B</b>	<b>Unicode Publications and Resources</b> .....	865
<b>B.1</b>	<b>The Unicode Consortium</b> .....	866
	The Unicode Technical Committee .....	866
	Other Activities .....	866
<b>B.2</b>	<b>Unicode Publications</b> .....	867
<b>B.3</b>	<b>Unicode Technical Standards</b> .....	868
	UTS #6: A Standard Compression Scheme for Unicode .....	868
	UTS #10: Unicode Collation Algorithm .....	868
	UTS #18: Unicode Regular Expressions .....	868
	UTS #22: Character Mapping Markup Language (CharMapML) .....	868
	UTS #35: Unicode Locale Data Markup Language (LDML) .....	868
	UTS #37: Unicode Ideographic Variation Database .....	868
	UTS #39: Unicode Security Mechanisms .....	868
	UTS #46: Unicode IDNA Compatibility Processing .....	869
<b>B.4</b>	<b>Unicode Technical Reports</b> .....	870
	UTR #16: UTF-EBCDIC .....	870
	UTR #17: Unicode Character Encoding Model .....	870
	UTR #20: Unicode in XML and Other Markup Languages .....	870
	UTR #23: The Unicode Character Property Model .....	870
	UTR #25: Unicode Support for Mathematics .....	870
	UTR #26: Compatibility Encoding Scheme for UTF-16: 8-Bit (CESU-8) ..	870
	UTR #33: Unicode Conformance Model .....	871
	UTR #36: Unicode Security Considerations .....	871
	UTR #50: Unicode Vertical Text Layout .....	871
	UTR #51: Unicode Emoji .....	871
<b>B.5</b>	<b>Unicode Technical Notes</b> .....	872
<b>B.6</b>	<b>Other Unicode Online Resources</b> .....	873
	Unicode Online Resources .....	873
	How to Contact the Unicode Consortium .....	874

<b>C</b>	<b>Relationship to ISO/IEC 10646</b> .....	875
<b>C.1</b>	<b>History</b> .....	876
<b>C.2</b>	<b>Encoding Forms in ISO/IEC 10646</b> .....	882
Zero Extending .....	882	
<b>C.3</b>	<b>UTF-8 and UTF-16</b> .....	883
UTF-8 .....	883	
UTF-16 .....	883	
<b>C.4</b>	<b>Synchronization of the Standards</b> .....	884
<b>C.5</b>	<b>Identification of Features for Unicode</b> .....	885
<b>C.6</b>	<b>Character Names</b> .....	886
<b>C.7</b>	<b>Character Functional Specifications</b> .....	887
<b>D</b>	<b>Changes from Previous Versions</b> .....	889
<b>D.1</b>	<b>Versions of the Unicode Standard</b> .....	890
<b>D.2</b>	<b>Clause and Definition Updates</b> .....	895
<b>E</b>	<b>Han Unification History</b> .....	897
<b>E.1</b>	<b>Development of the URO</b> .....	898
<b>E.2</b>	<b>Ideographic Rapporteur Group</b> .....	900
<b>E.3</b>	<b>CJK Sources</b> .....	902
<b>F</b>	<b>Documentation of CJK Strokes</b> .....	907
<b>R</b>	<b>References</b> .....	915
<b>R.1</b>	<b>Source Standards and Specifications</b> .....	916
<b>R.2</b>	<b>Source Dictionaries for Han</b> .....	925
<b>R.3</b>	<b>Other Script Sources</b> .....	926
<b>R.4</b>	<b>Selected Resources: Technical</b> .....	953
<b>R.5</b>	<b>Selected Resources: Other</b> .....	956
<b>I</b>	<b>General Index</b> .....	961

